Behind the scenes of submission processing at ClinVar

Melissa J Landrum, Shanmuga Chitipiralla, Garth Brown, Chao Chen, Baoshan Gu, Jennifer Hart, Douglas Hoffman, Wonhee Jang, Kuljeet Kaur, Chunlei Liu, Vitaly Lyoshin, Zenith Maddipatla, Rama Maiti, Joseph Mitchell, George Riley, Wenyao Shi, George Zhou, Donna R Maglott, J Bradley Holmes, Brandi L Kattman



ncbi.nlm.nih.gov 8



Upload submission to the Submission Portal

- ☐ Most data in ClinVar is submitted as a batch by file
- ☐ Single records can be submitted interactively with the ClinVar Submission Wizard
- ☐ Variants that must be updated quickly for a change in clinical significance can be updated with the Single SCV Update

What can you do to help us process your data faster?

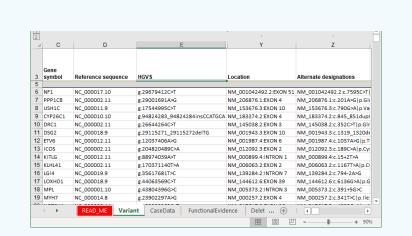
- ☐ Provide data in all required fields
- ☐ Validate your HGVS expression with Mutalyzer or VariantValidator
- ☐ Do not modify the column headers in the submission template
- ☐ Do not modify the cell validation; use allowed values when there is a list
- ☐ Check the instructions for each column on the spreadsheet for correct format and separators for a list



Review by curator and convert submission file to XML

Why do curators review each submission?

- Data can pass validation but not make biological sense
- Curators can ask for optional, but highly valued data, like the date the variant was last evaluated
- Curators help submitters correct errors and format data correctly



 ClinVar data is stored in XML, but many submitters don't use that format.

 Thus the first step for most submissions is to convert a different format - either a spreadsheet or a csv/tsv file to XML, with some validation.



Load submission database

SCV accession numbers

- ☐ Each submitted record is assigned an SCV accession number.
- ☐ The SCV accession is used to retrieve data and to provide updates for the record.

Loading to the database

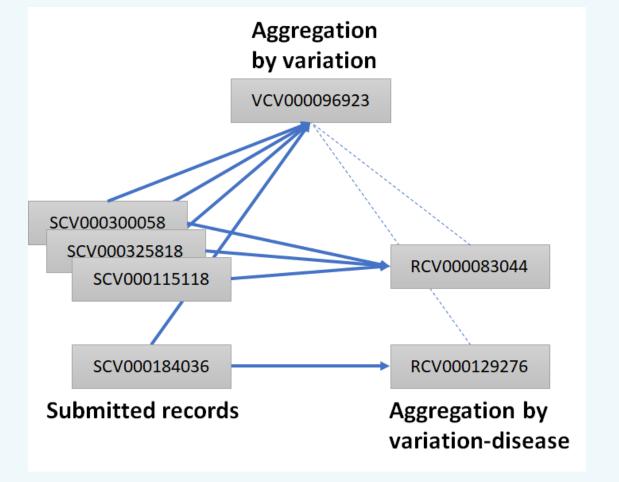
- ☐ Finds records submitted as novel that are truly updates
- ☐ Allows curators to review how variants and conditions are identified and aggregated
- ☐ Identifies errors in submitted data, such as duplicate records from the submitter or changes to the variant description in an update



Aggregate data and add annotation

ClinVar aggregates submitted data

- ☐ the VCV record: all data submitted for a variant
- ☐ the RCV record: all data submitted for a variantcondition pair

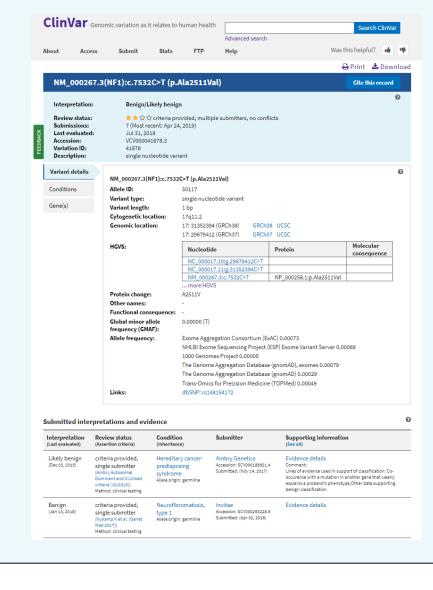


ClinVar adds annotation

- an identifier for the variant or set of variants (Variation ID)
- HGVS expressions on genomic DNA, cDNA, and protein
- genomic location on GRCh37 and GRCh38
- a predicted molecular consequence
- the gene(s) that colocalizes with each variant
- allele frequencies
- an aggregate interpretation, indicating consensus or conflict
- review status
- links to other databases

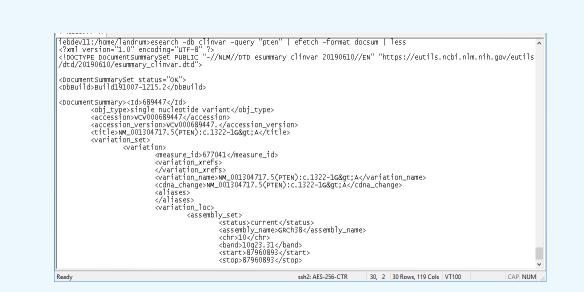


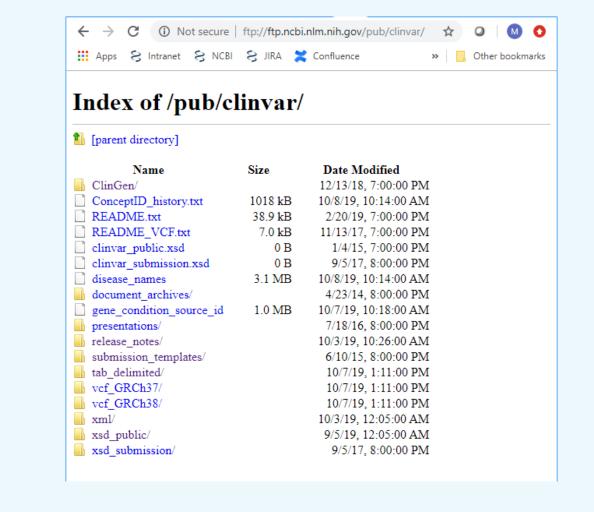
Submitted and aggregated data are published weekly



Submitted data are available

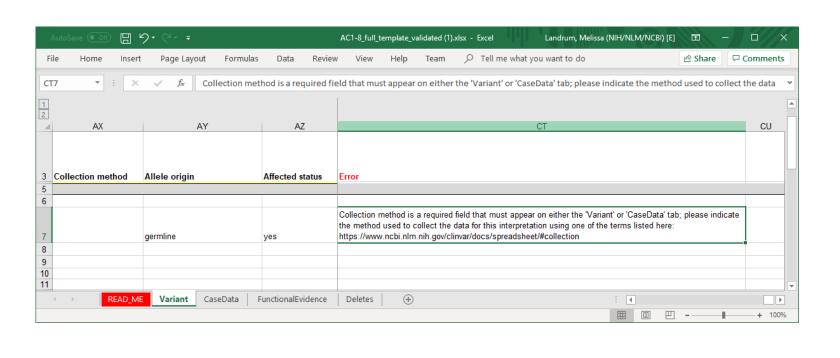
- ☐ On the ClinVar website
- ☐ Programmatically through E-utilities
- ☐ In XML, VCF, and tab-delimited files on the ftp site

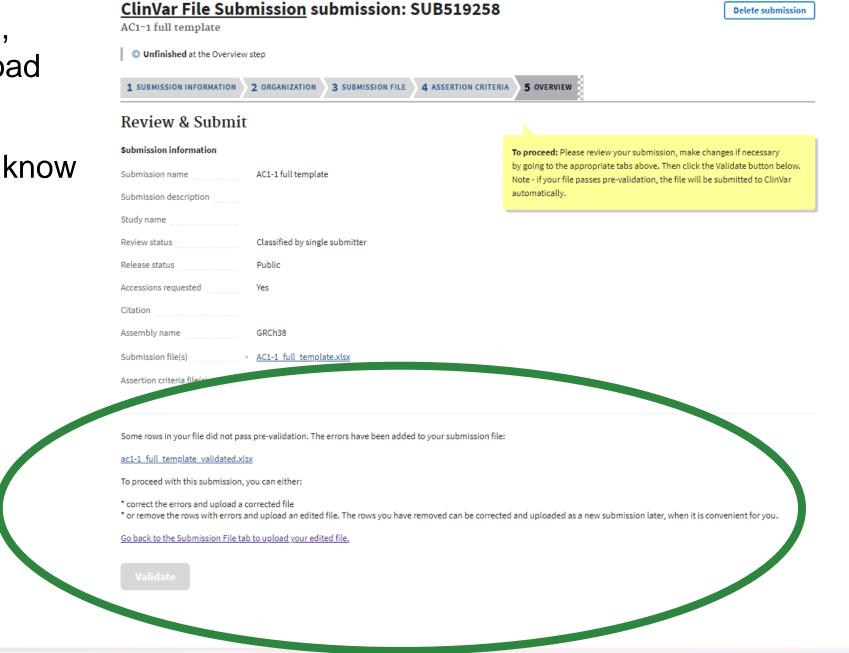




NEW!!! Pre-validation for ClinVar Submissions

- ☐ You now get some immediate feedback about clinical significance, collection method, allele origin, and affected status when you upload your file to the Submission Portal
- ☐ Error messages are added back to your submission file to let you know what the errors are and how to fix them.
- ☐ Checks for the variant and interpreted condition are coming soon!







Visit NCBI staff at Booth 214!



Acknowledgements

This research was supported by the Intramural Research Program of the NIH, National Library of Medicine.

Need to reach us? Go to support.nlm.nih.gov.



ncbi.nlm







